

How to use Corpora in Translation Work

Vellano, Olaszország – 1999. június

Pápai Vilma

Ha azt halljuk: 200 milliós szövegbank, Európai Unió támogatás, korpusznyelvészet, számítógépes laboratórium, online korpuszelérhetőség, minden bizonynyal zajos nagyvárosi környezetre, pazar irodákra gondolunk. Ha pedig azt halljuk: napernyők alatt, csendes, meghitt beszélgetés közben elköltött ebéd, erdei séták, vacsora a falusi panzió vendéglőjében, akkor a nagyvilágtól távoli vidék képe jelenik meg előttünk. Hogy mindezen kellékek együtt is jelen lehetnek, arra jó példa a Toszkánában, a Firenze fölötti hegyek között fekvő kis Vellano falu, ahol 1999. június 21. és 27. között a fenti címmel rendezett tanfolyamot a Tuscan Word Centre (TWC).

Ebben az elzárt faluban, amelyet tömegközlekedési eszközzel szinte lehetetlen megközelíteni, egy héti intenzív eszmecsere folyt a nyelvvel foglalkozó és a korpusznyelvészet iránt érdeklődő résztvevők, valamint a tudományágat képviselő szervezők között.

S a szervezők nem voltak mások, mint a nyelvészet egyre inkább megerősödő fiatal ágának, a korpusznyelvészetnek nemzetközi hírű művelői: *John Sinclair*, a tudományterület egyik vezető tudósa, a birminghami egyetem professzora, számos brit és európai szintű lexikográfiai, valamint gépi fordítással foglalkozó projekt vezetője és *Elena Tognini Bonelli*, jelenleg az olaszországi Lecce egyetemének professzora. További előadónak megnyerték *Mona Bakert*, a fordításelmélet és a korpusznyelvészet egyaránt elismert képviselőjét, aki szerteágazó tudományos munkássága mellett szerkeszti a *The Translator* című folyóiratot és vezeti a manchesteri egyetem, a UMIST (University of Manchester, Institute of Science and Tech-

nology) alkalmazott nyelvészeti tan-
székét.

Bár a kurzus 20 résztvevője különböző területekről: a nyelvészet, a nyelvoktatás, a fordítástudomány, a nemzeti korpuszépítés, az informatika, sőt a könyvkiadás területéről s – nem mellékesen – 11 európai országból érkezett, az egy-két főből álló nemzeti csoportok egy jövőbeli Európai Unió szellemében kialakított, soknemzetiségű, de szakmailag homogén munkaközösséget alkottak. Magyarországot Balaskó Máriaival (Berzsenyi Dániel Tanárképző Főiskola) ketten képviseltük.

A Tuscan Word Centre szellemes elnevezés mögött az Európai Unió támogatásával a nyelvtudományi kutatások elősegítésére létrehozott non-profit szervezet húzódik meg, amelynek működéséhez – s különösen egyes szemináriumainak anyagi támogatásához – az olaszországi British Council is hozzájárul. Rövid intenzív tanfolyamok szervezésén kívül tanácsadó és szakértői tevékenységet is végez, konferenciák, szemináriumok, projektek tervezését és lebonyolítását, továbbá nyelvfeldolgozó szoftver minősítését is vállalja.

A TWC tevékenysége alapvetően a korpusznyelvészethez kapcsolódik. A korpusznyelvészet felismerte azt a napjainkban még felmérhetetlen lehetőséget, ami a számítógépen tárolt és számítógéppel kezelhető nagy szövegbankokban rejlik. A nagy terjedelmű szövegek olyan forrást jelentenek, amelyek új szemléletű nyelvelírást tesznek lehetővé. Nevezetesen: a korpusz nyújtotta adatokból autentikusabb és árnyaltabb tudást nyerhetünk a nyelvről, mint a nyelvtanítók intuícióján alapuló példamondatokból vagy a szótárakban a jelentés illusztrációjára hozott, ugyancsak többnyire egyetlen nyelvhasználó által megalkotott szótári példamondatokból. A folyamatosan bővített korpuszok másik nagy előnye, hogy a nyelv aktuális állapotát mutatják, következőképpen érzékenyen tükrözik a nyelvi változásokat is.

A korpuszokból nyert és helyesen értelmezett információ segítségével hatékonyabb lesz mindennapi és szakmai kommunikációnk is, legyen szó jogi, kereskedelmi, politikai vagy egyéb szakterületen végzett kommunikációról. Nem véletlen, hogy az Európai Unió már számos projektet (például a PAROLE, TELRI és ELAN) indított többnyelvű vizsgálatok folytatására és többnyelvű adatbankok létrehozására.

A vizsgálatok egyik fő célja a *fordítási folyamat automatizálásának* meggyorsítása, és ezáltal – többek között – az Európai Unióra háruló hatalmas fordítási feladat megkönnyítése. Ez volt az elsődleges célja az 1990-ben elkezdett többnyelvű lexicográfiai projektnek (Council of Europe Multilingual Lexicography Project) is, amelyben John Sinclair és a birminghami egyetem korpusznyelvészeti csoportja vezetésével számos európai egyetem kutatói vettek részt. A vizsgálatban, amelyben hét európai nyelv között a magyar is szerepelt (angol, horvát, magyar, német, olasz, spanyol és svéd), arra a kérdésre kerestek választ, hogy milyen megfelelés áll fenn *A* nyelv egy-egy szavának szöveggörnyezete és *B, C, D* stb. nyelvi ekvivalensének szöveggörnyezete között, és hogyan írható le ez a megfelelés a számítógép számára.

Az ez évi júniusi TWC-tanfolyamon a szöveggörnyezetek felhasználásának másik lehetséges területe állt a vizsgálódás középpontjában: hogyan alkalmazhatók a *korpuszok a fordítás elméletében és gyakorlatában*.

A tanfolyamon naponta hat óra előadást hallgattunk és négy óra gyakorlati foglalkozáson vettünk részt. Az előadások sorát megnyitó Sinclair professzor a korpuszok felhasználásának általános elméleti és módszertani kérdéseiről beszélt. Az előbbi területhez tartoznak például a korpuszok és a nyelvészet kapcsolatáról, a korpuszok típusairól, a tervezésüket vezérlő elvekről, külső és belső kritériumokról szóló előadások. A metodikai kérdé-

sek közül érintette a lekérdező programok témakörét, a korpuszba felvett anyag feldolgozásának, pontosabban feldolgozottságának különbségeit. A kutatás céljától függ, hogy ún. nyers, folyamatos szövegeket gyűjtünk vagy részletes kódrendszerrel ellátott, elemzett korpuszt építünk. Tanácsokat kaptunk olyan problematikus kérdésekben, hogyan lehet a korpuszvizsgálatkor helyes kérdéseket feltenni, hogyan kell a kapott adatokat értelmezni, értékelní és nyelvészeti következtetéseket levonni.

Az elméleti és metodikai kérdések tárgyalása nem különült el a gyakorlattól, ezeket a professzor azonnal gyakorlati példákkal mutatta be. A birminghami központra csatlakozva az angol Cobuild-korpusz, a Bank of English mintegy 200 millió szóból álló készletét használta fel illusztrációként. Számos példán mutatta be, hogyan lehet a korpusz segítségével egy-egy szóról vagy szókapcsolatról nyelvi információt szerezni, pontosabban szólva, hogyan lehet a vizsgált nyelvi egység *tipikus* viselkedését feltárni a kontextus alapján.

A második előadó, Mona Baker kutatói érdeklődése elsősorban a fordított szövegek sajátosságaira irányul. Vizsgálataihoz egy másik korpuszt használ, az úgynevezett TEC-korpuszt (Translational English Corpus). A TEC 5 millió szavas gyűjteménye kizárólag fordítás eredményeként létrejött angol szövegeket tartalmaz, amelyben minden szöveget a manchesteri kutatócsoport igen részletes bibliográfiai információval látott el, például: a fordító nemzetisége, neme, alkalmazási státusa, a forrásnyelvi szöveg, a szerző és a kiadó(k) adatai, a fordítás folyamat jellemzői.

Vitát váltott ki az a kérdés, hogy milyen fordított szövegeket kell felvenni a korpuszba: csak lektorált és professzionális fordító által készített fordításokat vagy kevésbé gyakorlott fordítók munkáit is. Egy reprezentatívnek nevezett gyűjteménynek ugyanis az utóbbiakat is tartal-

maznia kell. Ez a szembeállítás visszavezet bennünket a fordításelmélet klasszikus kérdéséhez: a deskriptív vagy normatív megközelítés szembeállításához. A TEC-ben jelenleg csak nyomtatásban megjelent, a fordító anyanyelvére fordított művek szerepelnek.

Mona Baker bevezetőjében szervesen kapcsolódott Sinclair professzor előadásaihoz. Röviden ismertette, hogy a Bank of English gyűjteményében kimutatott törvényszerűségek kiállják-e a próbát a fordított szövegek korpuszában. Megvizsgálta tehát, hogy az angol nemzeti korpuszból nyert kollokációs mintákat tükrözik-e a fordított szövegek is. Minden alkalommal meg kellett állapítania, hogy a gyors vizsgálat alapján a fordítás eredményeként létrejött angol szövegek nem mutatják azokat a szabályszerűségeket, mint az eredeti szövegek.

Előadásaiban beszélt továbbá a fordítástudományban és -oktatásban felhasználható korpuszok fajtáiról: a kétnyelvű vagy párhuzamos, a többnyelvű és az ún. összehasonlítható egynyelvű korpuszokról. A párhuzamos korpusz egy adott nyelven írt műből és adott célnyelvi fordításából épül fel. A többnyelvű korpuszok (multilingual corpora) olyan két vagy több egynyelvű korpuszt tartalmaznak, amelyek egymásnak nem fordításai, mint például a fentebb említett Council of Europe Multilingual Project korpusza. Összevetésük mégis lehetséges, mert hasonló kritériumok alapján hozták létre őket. Az ún. összehasonlítható (comparable corpora) korpuszok két különböző korpuszból állnak, amelyek azonos nyelvű szövegeket tartalmaznak: míg az egyik az adott nyelven írt eredeti szövegekből, a másik az adott nyelvre fordított szövegekből áll. Mona Baker szerint az összehasonlítható korpuszok tanulmányozásával eljuthatunk azokhoz a tulajdonságokhoz, amelyek minden fordított szövegre jellemzőek. A korpuszok segítségével alátámaszthatjuk vagy megcáfolhatjuk a fordításelméletben élő

hipotéziseket. Példaként említhetjük az explicitációs hipotézist vagy Gideon Toury alábbi állítását is: "in translation, textual relations obtaining in the original are often modified, sometimes to the point of being totally ignored, in favour of (more) habitual options offered by a target repertoire" (Toury 1995: 268).

A különböző korpusztípusok fordítás-tudományi felhasználásáról szólva Mona Baker felhívta figyelmünket több lehetséges kutatási témára is. Nem hallgatta el azonban az egyes korpuszfajták alkalmazásának nehézségeit sem. Bár léteznek kereskedelemben kapható programok, például a párhuzamos szövegek létrehozásához szükséges ún. illesztő programok, mégis az adott kutatás igényei szerinti módosításukhoz, illetve működtetésükhöz jelentős konzultációra van szükség számítógépes szakemberekkel. Magam számára éppen ezt a következtetést vontam le, hogy a korpuszok gyakorlati alkalmazásához jelentős team-munka szükséges, elsősorban a számítástechnikai háttér biztosítása érdekében.

Elena Tognini Bonelli előadásai voltak a leginkább gyakorlat központúak. Újabb kutatási témáit tárta elénk, példákat hozva a korpuszok gyakorlati felhasználásának lehetőségeire. Nagy érdeklődést keltettek az egy-egy szinonima-pár kollokációinak elkülönítésére folytatott vizsgálatai.

A gyakorlati foglalkozásokon kis csoportokban végeztünk korpuszkutatási feladatokat. A feladatok lépésről-lépésre vezettek végig egy-egy szókapcsolat szöveggörnyezete törvényszerűségeinek feltárásában a következtetések levonásáig. Tanulságosak voltak a résztvevők kutatómunkájáról szóló beszámolók is, amelyek bizonyították, hogy a korpuszépítés valóban jelen van és népszerű Európa egyetemlein. Ahogy Sinclair professzor megjósolta:

"Thirty years ago when this research started it was considered impossible to process texts of several million

words in length. Twenty years ago it was considered marginally possible but lunatic. Ten years ago it was considered quite possible but still lunatic. Today it is very popular". (Sinclair 1991:1)

A Toszkanában töltött egy hét alatt a napi tíz óra munkát kötetlen beszélgetések egészítették ki. Bepillanthattunk a lelkes kutatógárda alkotóműhelyébe, igényes, tanítványokat is nevelő munkájába. Mi több, a két tudományág e kiemelkedő egyéniségeiben szívélyes, önzetlen és nagylelkű embereket ismerhettünk meg. Így a szakmai útravaló mellé ajándékot is kaptunk. Ennek az élménynek a nagyszerűségét.

Irodalom

- Toury, G. 1995. *Descriptive Translation Studies and beyond*. Amsterdam: Benjamins.
Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Bradfordi nyári egyetem tolmácsoknak

Bradford, Anglia, 1999. június 21 – július 16.

Horváth Ildikó

Az Európai Parlament felkérésére a Bradfordi Egyetem Modern Nyelvek Tanszéke 1999. június 21. és július 16. között első ízben szervezett négyhetes nyári kurzust fiatal, frissen végzett tolmácsok számára. A tanfolyamra 5 közép- és kelet-európai országból összesen 25 hallgató (5 lengyel, 4 cseh, 4 észt, 4 lett, 4 szlovén és 4 magyar) érkezett. A kurzus célja a résztvevők angoltudásának intenzív fejlesztése volt.

Ennek érdekében a házigazdák minden tőlük telhetőt megtettek. A kurzus ideje alatt összesen 36 oktató töltött több-kevesebb időt a hallgatókkal: volt, akivel

minden nap találkoztunk, de egyes oktatók csak egy-két alkalommal tartottak előadást. A házigazda szerepét az egyetem tolmácsképző csoportja töltötte be, melynek frissen végzett hallgatói közül többen folyamatosan a külföldiek rendelkezésére álltak. Az ő társaságukban számos vidám és nyelviileg hasznos órát töltöttünk el. A tanfolyamot a tolmácsképző csoport termeiben tartották és az intézmény könyvtárát/mediotékáját is természetesen használhattuk. Ezenkívül szabad bejárásunk volt az egyetem többemeletes könyvtárába és az uszodába is.

A kurzus műszaki felszereltségére az jellemző, hogy minden hallgató minden nap megkapta valamelyik angol napilap aktuális számát, ezen kívül minden nyelvcsoporthoz elláttak a *Financial Times*-szal, a heti rádió és tv műsorrall, valamint egy másik napilappal. Mindenkinek rendelkezésére bocsátottak egy rádiós magnót, továbbá két-két hallgató kölcsön kapott egy-egy televíziót videomagnóval.

Az oktatás reggel 8.30-kor kezdődött és általában 17.30-kor fejeződött be. Az esti programokra csak ezt követően került sor, melyek látogatása (közös indiai vacsorák, filmvetítés, pub-látogatás) többnyire fakultatív volt. A reggeli órát a média (sajtó, rádió, tv) tanulmányozásának szentelték a szervezők. Ez lehetőséget adott arra, hogy Nagy-Britannia aktuális politikai és a közvéleményt éppen foglalkoztató eseményeibe betekintést nyerjünk. A formális nyelvórákra csak ezt követően került sor. Órarendünkben a hagyományos nyelvórákon kívül, melyek elsősorban a szókincsfejlesztésre koncentráltak, a következő foglalkozások szerepeltek: előadásmód, tolmácsolás (konszekutív és szinkron), ízelítő az angol nyelv különböző dialektusaiból, az angol, mint világnyelv. A meghívott előadók nagyon változatos témákban tartottak előadást. Halottunk többek között a diplomáciai és a törvényeségi tolmácsolásról, az EU különböző intézményeinek konferenciatolmácsolási